255

Reihe Ökonomie

Economics Series

Spatial Chow-Lin Methods for Data Completion in Econometric Flow Models

Wolfgang Polasek, Richard Sellner

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

255

Reihe Ökonomie

Economics Series

Spatial Chow-Lin Methods for Data Completion in Econometric Flow Models

Wolfgang Polasek, Richard Sellner

September 2010

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Wolfgang Polasek
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-155
email: polasek@ihs.ac.at
and
University of Porto
Faculty of Science
Rua Dr. Roberto Frias
4200 Porto, Portugal

Richard Sellner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-261
email: sellner@ihs.ac.at

Abstract

Flow data across regions can be modeled by spatial econometric models, see LeSage and Pace (2009). Recently, regional studies became interested in the aggregation and disaggregation of flow models, because trade data cannot be obtained at a disaggregated level but data are published on an aggregate level. Furthermore, missing data in disaggregated flow models occur quite often since detailed measurements are often not possible at all observation points in time and space. In this paper we develop classical and Bayesian methods to complete flow data. The Chow and Lin (1971) method was developed for completing disaggregated incomplete time series data. We will extend this method in a general framework to spatially correlated flow data using the cross-sectional Chow-Lin method of Polasek et al. (2009). The missing disaggregated data can be obtained either by feasible GLS prediction or by a Bayesian (posterior) predictive density.

Keywords

Missing values in spatial econometrics, MCMC, non-spatial Chow-Lin (CL) and spatial Chow-Lin (SCL) methods, spatial internal flow (SIF) models, origin and destination (OD) data

JEL Classification

C11, C15, C52, E17, R12

Comments
This paper is part of a project funded by the Jubilaeumsfonds of the Austrian National Bank (OeNB).

Contents

1. Introduction

The origin of the Chow-Lin method lies in the desire to complete data sets for disaggregated time series problems. This paper will do an extension in two directions: First, we will use a spatial econometrics model and then we will use it for classical and Bayesian estimation for origin and destination (OD) data in flow models as in LeSage and Pace (2008). We propose a spatial econometrics model in a Bayesian framework that will be estimated by MCMC.

The name internal flows stems from the fact that we consider flows between n disaggregate units that will be aggregated to N aggregate units. Such model we call spatial internal flows (SIF) models because we aggregate data for flows within a fixed geographic area like a country.

This paper derives spatial Chow-Lin methods for flow data matrices, based on models for origin to destination (OD) flows (see LeSage and Pace (2008)) and explains the GLS and the feasible GLS approach and the Bayesian approach to estimate general SIF models. Simplified SIF models are called uni-lateral or origin (oSIF) or destination (dSIF) models because they concentrate only on one variance component of the spatial correlation polynomial. For the Bayesian treatment of these SIF models we have to elicit a prior distribution and then we explain how to adopt a heteroskedastic MCMC algorithm for estimation.

The spatial modeling of flow models face the problem that large cross sections imply rather large spatial weight matrices which makes any estimation procedures computationally expensive. The plan of the paper is as follows. In the next section we describe the basic spatial internal flow (SIF) model. Then we derive the Chow-Lin procedure for non-spatial flow models, before we explain the Chow-Lin procedure for spatial flow models. Finally, we apply the model to European trade flow data. In a final section we conclude.

2. Completing data in spatial internal flow (SIF) models

We adopt the following notation: Let $\mathbf{Y}_a : N \times N$ be the aggregated flow matrix for N aggregated cross-sectional units and $\mathbf{Y}_d : n \times n$ be the disaggregated panel matrix. For a flow matrix the aggregation has to be done in 2 dimensions:

$$\mathbf{Y}_a = \mathbf{C}_0 \mathbf{Y}_d \mathbf{C}_0'. \tag{1}$$

The aggregation matrix is $\mathbf{C}_0 : N \times n$ with $n > N$ across spatial units has to be defined as a block diagonal matrix (as in Polasek et al. (2009)):

$$\mathbf{C}_0 = diag(\mathbf{1}_{n_1}', ..., \mathbf{1}_{n_N}'), \quad \sum_{i=1}^{N} n_i = n, \tag{2}$$

where the $n_i's$ is the number of sub-units to be aggregated in each cell (unit) and $\mathbf{1}_{n_i} : n_i \times 1$ is a column vector of ones and indexes the areas where units are aggregated. $\mathbf{Y}_d$ is the $n \times n$ disaggregated matrix. The sub-lengths add to the total number: $n_1 + ... + n_N = n$.

For the Chow-Lin procedure we have to vectorize the aggregation equation of the flows:

$$\mathbf{y}_a = (\mathbf{C}_0 \otimes \mathbf{C}_0)\mathbf{y}_d = \mathbf{C}\mathbf{y}_d \quad \text{with} \quad \mathbf{C} = \mathbf{C}_0 \otimes \mathbf{C}_0, \tag{3}$$

with $\mathbf{y}_d = vec\mathbf{Y}_d : n^2 \times 1$ and the joint aggregation matrix is $\mathbf{C} = \mathbf{C}_0 \otimes \mathbf{C}_0 : N^2 \times n^2$, because of $vec(ABC) = (C' \otimes A)vecB$.

We need a fully observed disaggregated panel matrix $\mathbf{X}_d : n \times K$ (the aggregated matrix is $\mathbf{X}_a : N \times K$) as "panel indicators", which can be vectorized to a $nK \times 1$ vector $vec\mathbf{X}_d = \mathbf{x}_d$. Note that indicator matrices for the disaggregated flows need to have the same dimension as $\mathbf{Y}_d$ (and the same number K). The disaggregated model is a linear regression model using the vectorized flow matrices:

$$\mathbf{y}_d = \mathbf{X}_d\boldsymbol{\beta}_d + \mathbf{u}_d, \quad \mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \Omega_d \otimes \sigma^2\mathbf{V}_d], \tag{4}$$

where $\mathbf{V}_d$ and $\boldsymbol{\Omega}_d$ are the disaggregated $n \times n$ covariance matrices, and $\mathbf{y}_d = vec\mathbf{Y}_d$ is the vectorization of the flow matrix $\mathbf{Y}_d$. The covariance matrices have the following interpretation: Ω_d is the covariance matrix across (between) the columns while $\mathbf{V}_d$ is the covariance matrix within the columns. A simpler assumption for the covariance matrix is the assumption of homoskedasticity (and uncorrelatedness):

$$\mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2\mathbf{I}_{nn}] \quad \text{with} \quad \mathbf{I}_{nn} = \mathbf{I}_n \otimes \mathbf{I}_n. \tag{5}$$

For such a model we have to vectorize the flow matrix $\mathbf{Y}_d$ and we use as indicators in the regression distances and the origin and destination variables.

Definition 1 (The SIF model). *We consider the disaggregated dependent variable $\mathbf{y}_d = vec\mathbf{Y}_d$ of a flow matrix $\mathbf{Y}_d : n \times n$ and we assume a SAR model of the form as in LeSage and Pace (2008). Such a regression model we will call a SIF (spatial internal flow) model for flow (or origin-destination) data:*

$$\mathbf{y}_d = \rho(\mathbf{W}_1, \mathbf{W}_2)\mathbf{y}_d + \mathbf{X}_d\beta_d + \mathbf{u}_d, \tag{6}$$

where $\rho(\mathbf{W}_1, \mathbf{W}_2)\mathbf{y}_d$ stands for a spatial lag polynomial that captures spatial correlation structures and is applicable for flow models

$$\rho(\mathbf{W}_1, \mathbf{W}_2)\mathbf{y}_d = \rho_1(\mathbf{W}_1 \otimes I_n)\mathbf{y}_d + \rho_2(\mathbf{I}_n \otimes \mathbf{W}_2)\mathbf{y}_d + \rho_3(\mathbf{W}_1 \otimes \mathbf{W}_2)\mathbf{y}_d. \tag{7}$$

The SIF model is homoskedastic if the residuals are distributed as $\mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2\mathbf{I}_{nn}]$ and heteroskedsatic if the residuals are distributed as $\mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2\Omega_d \otimes \mathbf{V}_d]$.

The spatial correlation is decomposed into 3 components: ρ_1 is attributed to the spatial correlation of the rows (destination sites), ρ_2 to the column (origin) component and ρ_3 is the interaction component. Simpler "unilateral" SIF models can be obtained if we consider just 1 component of the rho-polynomial (see section 4 for more on origin and destination components). The aggregated reduced form (ARF) of the SIF model (6) is given by multiplying the reduced form by the aggregation matrix $\mathbf{C}$:

$$\mathbf{C}\mathbf{y}_d = \mathbf{C}\mathbf{R}_\rho^{-1}\mathbf{X}_d\beta_d + \mathbf{C}\mathbf{R}_\rho^{-1}u_d, \tag{8}$$

where the general spread matrix $\mathbf{R}_\rho$ for spatial flow models is given by

$$\mathbf{R}_\rho = \mathbf{I}_{nn} - \rho_1(\mathbf{W}_1 \otimes \mathbf{I}_n) + \rho_2(\mathbf{I}_m \otimes \mathbf{W}_2) + \rho_3(\mathbf{W}_1 \otimes \mathbf{W}_2), \tag{9}$$

and $\mathbf{W}_1$ and $\mathbf{W}_2$ are suitable chosen neighborhood matrices (see Anselin (1988) or LeSage and Pace (2009) for discussion on possible $\mathbf{W}$'s).

The reduced form (RF) of the spatial internal flow (SIF) model is obtained by collecting all the dependent variables on the left hand side

$$\mathbf{y}_d = \mathbf{R}_\rho^{-1}\mathbf{X}_d\beta_d + \tilde{\mathbf{u}}_d, \quad \tilde{\mathbf{u}}_d = \mathbf{R}_\rho^{-1}\mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2\mathbf{V}_\rho], \tag{10}$$

and the reduced form variance covariance matrix (VCV) is a function of the unknown parameter ρ:

$$\mathbf{V}_\rho = \mathbf{R}_\rho^{-1}(\Omega_d \otimes \mathbf{V}_d)\mathbf{R}_\rho^{'-1}. \tag{11}$$

Because the disaggregated model can not be used to estimate the disaggregated model parameters $\theta_d = (\beta_d, \rho_d, \sigma_d^2)$, we transform the model in order to get a fully observed data set. If the aggregate data are known we can transform (= aggregate) the disaggregated data to an estimable equation with aggregated data using $\mathbf{y}_a = \mathbf{C}\mathbf{y}_d$. Now the estimation can be done using the aggregated reduced form ARF of the SIF model in (8).

Since only the aggregated data are completely observed we have to make a connection between the aggregated model and the disaggregated model and we adopt a notation that can separate the 2 models. In compact notation the spatial ARF model is obtained through the aggregation matrix $\mathbf{C} : N^2 \times n^2$ in (3):

$$\mathbf{y}_a = X_{a\rho}\beta_d + u_{a\rho}, \quad u_{a\rho} \sim \mathcal{N}[\mathbf{0}, \sigma^2 V_{a\rho}], \tag{12}$$

where $\mathbf{u}_{a\rho}$ is the aggregated residual and the covariance matrix is

$$\mathbf{V}_{a\rho} = \mathbf{C}\mathbf{R}_\rho^{-1}(\Omega_d \otimes \mathbf{V}_d)\mathbf{R}_\rho^{\prime -1}\mathbf{C}'. \tag{13}$$

There exist useful relationships between the aggregated and disaggregated variables: $\mathbf{y}_a = \mathbf{C}\mathbf{y}_d$ is the direct aggregation for the dependent variable, but – interestingly – the regressor variables follow an indirect aggregation rule: $X_{a\rho} = \mathbf{C}\mathbf{R}_\rho^{-1}\mathbf{X}_d$, because of the inverse spatial correlation matrix $\mathbf{R}_\rho$ sitting between the aggregation matrix and the disaggregated observations.

3. Non-spatial internal flow (nSIF) models

Since the econometric analysis of flow models involves high dimensions and is more demanding, it is useful to start explaining the modeling process of flows in a non-spatial model. It will help to understand the extension to the spatial modeling process.

Spatial internal flow (SIF) models are high dimensional models that grow with the square of the number of cross sections. The spatial lag assumption introduces a spatial filter that makes the model non-linear in the spatial correlation parameter and creates regressors and covariance matrices that disturbs the otherwise nice Kronecker structure of the flow model. Therefore we like first to see how the 'spatial warping' of the variables (through the spread matrix $\mathbf{R}$) or the "spatial curse" of dimensionality can be avoided by estimating a non-spatial internal flow ('nSIF') model.

Definition 2 (Non-spatial flow (nSIF) models). *The (heteroskedastic) non-spatial internal flow (nSIF) model for disaggregated data is given in matrix form by*

$$\mathbf{Y}_d = \sum_{i=1}^{K} \mathbf{X}_{di}\beta_{di} + \mathbf{U}_d, \quad \mathbf{U}_d \sim \mathcal{N}_{n\times n}[\mathbf{0}, \sigma_d^2 \Omega_d \otimes \mathbf{V}_d], \tag{14}$$

where K is the number of regressors and $\mathcal{N}_{n\times n}$ denotes the matrix normal distribution, and there are K disaggregated regressor panels $\mathbf{X}_{di} : n \times n, i = 1, ..., K$.
The (homoskedastic) non-spatial internal flow (nSIF) model makes the following simplified assumption for the error structure:

$$\mathbf{U}_d \sim \mathcal{N}_{n\times n}[\mathbf{0}, \sigma_d^2 \mathbf{I}_{nn}]. \tag{15}$$

In contrast, the SIF model in matrix form for aggregated data has the form

$$Y_a = \sum_{i=1}^{K} \mathbf{X}_{ai}\beta_{ai} + U_a, \quad U_a \sim \mathcal{N}_{n\times n}[0, \sigma_a^2 \Omega_{agg} \otimes \mathbf{V}_{agg}], \tag{16}$$

where the aggregated data are $Y_a = CY_dC' : N \times N$ and the scalar coefficient β_{ai} is the i-th element of the regression vector $\beta_a : K \times 1$ in the aggregated model. For this model we obtain different residuals and residual covariance matrices $\Omega_{agg} \otimes \mathbf{V}_{agg}$.

3.1. Least squares (LS) estimation for the non-spatial internal flow (nSIF) model

For the non-spatial model (16) we consider various estimation procedures, the first one being the least squares approach. Assume that there are K panel indicator matrices $\mathbf{X}_{d1}, \ldots, \mathbf{X}_{dK}$ available for the disaggregate model and the first one $\mathbf{X}_{d1}$ defines the regression constant by a matrix of ones $\mathbf{X}_{d1} = \mathbf{1}_n \otimes \mathbf{1}_n'$. Furthermore, we define the regressor matrix of all vectorized panel regressors

$$\tilde{\mathbf{X}}_d = (vec\mathbf{X}_{d1}, \ldots, vec\mathbf{X}_{dK}) : (nn \times K) \quad and \quad \mathbf{C}\tilde{\mathbf{X}}_d = \tilde{\mathbf{X}}_a : (NN \times K). \tag{17}$$

The aggregated model is obtained by multiplying the regression equation with the aggregation matrix $\mathbf{C}$ as in (14) and we obtain

$$\begin{aligned} \tilde{\mathbf{X}}_a &= (vec(\mathbf{C}\mathbf{1}_n\mathbf{1}_n'\mathbf{C}'), vec(\mathbf{C}\mathbf{X}_{d2}\mathbf{C}'), ..., vec(\mathbf{C}\mathbf{X}_{dK}\mathbf{C}')) = \\ &= (vec\mathbf{X}_{a1}, vec\mathbf{X}_{a2}, ..., vec\mathbf{X}_{aK}). \end{aligned} \tag{18}$$

Note the relationship between the K disaggregated and the aggregated indicator matrices: $\mathbf{X}_{dk} : (n \times n) \to \mathbf{X}_{ak} : (N \times N), k = 1, \ldots, K$ which have to be vectorized to build up the regressor matrix. The transposed regressor matrix $\tilde{\mathbf{X}}_a$' is given by

$$\tilde{\mathbf{X}}_a' = \begin{pmatrix} vec'\mathbf{X}_{a_1} \\ .. \\ vec'\mathbf{X}_{a_K} \end{pmatrix} : (K \times N^2), \tag{19}$$

since there are N^2 elements per row and the aggregated model can be written as

$$\mathbf{y}_a = \tilde{\mathbf{X}}_a\beta_a + \mathbf{u}_a. \tag{20}$$

To estimate the covariance matrices Ω_a and $\mathbf{V}_a$, we first estimate β_a by OLS ($\beta_a^{OLS} : (K \times 1)$), using using the vectorized panel matrices in (18):

$$\beta_a^{OLS} = (\tilde{\mathbf{X}}_a'\tilde{\mathbf{X}}_a)^{-1}\tilde{\mathbf{X}}_a'\mathbf{y}_a. \tag{21}$$

Construct the residual matrix $\hat{\mathbf{U}}_a$ from the OLS residuals $\hat{u}_a = \mathbf{y}_a - \tilde{\mathbf{X}}_a\beta_a^{OLS}$ then we get the covariance estimates

$$\hat{\mathbf{\Omega}}_a = \hat{\mathbf{U}}_a'\hat{\mathbf{U}}_a/N \quad \text{and} \quad \hat{\mathbf{V}}_a = \hat{\mathbf{U}}_a\hat{\mathbf{U}}_a'/N. \tag{22}$$

With these sample covariance matrices from the aggregated model we can estimate $\hat{\Sigma}_a = \hat{\Omega}_a \otimes \hat{\mathbf{V}}_a$ and obtain the feasible GLS estimate for the aggregated level model

$$\beta_a^{FGLS} = (\mathbf{X}_a'\hat{\Sigma}_a^{-1}\mathbf{X}_a)^{-1}\mathbf{X}_a'\hat{\Sigma}_a^{-1}\mathbf{y}_a. \tag{23}$$

Because the vectorization of the flow matrices leads to high dimensions of the involved matrices, we show in the next theorem how to simplify the moment matrices of the GLS estimator.

Theorem 1 (Simplified moment matrices for GLS and FGLS). *The GLS estimate for β_d in the aggregated SIF model (16) can be found by the $K \times 1$ estimator using estimates of moments*

$$\beta_d^{GLS} = \mathbf{M}_{\tilde{X}\tilde{X}}^{-1}\mathbf{M}_{\tilde{X}\tilde{Y}}, \tag{24}$$

and the feasible GLS estimator is

$$\beta_{FGLS} = \hat{\mathbf{M}}_{\tilde{X}\tilde{X}}^{-1}\hat{\mathbf{M}}_{\tilde{X}\tilde{Y}}, \tag{25}$$

where $\hat{\mathbf{M}}_{\tilde{X}\tilde{X}}$ *and* $\hat{\mathbf{M}}_{\tilde{X}\tilde{Y}}$ *denote the estimated moment matrices and* V_a *is replaced by a point estimate. The tilde indicates that we replace the theoretical covariance matrices by the estimated ones* $\hat{\Omega}_d$ *and* $\hat{V}_d$.

Proof 1. *The aggregate regressor moment matrix* $\mathbf{X}_a'\mathbf{V}_{a\rho}\mathbf{X}_a$ *for the GLS estimation is*

$$\mathbf{M}_{\tilde{X}\tilde{X}} = \left(\begin{pmatrix} vec'\mathbf{X}_{a_1} \\ \ldots \\ vec'\mathbf{X}_{a_K} \end{pmatrix} (\Omega_a \otimes V_a)^{-1}\mathbf{X}_a \right) =$$

$$\begin{pmatrix} trV_a^{-1}\mathbf{X}_{a1}\Omega_a^{-1}\mathbf{X}_{a1}' & \ldots & trV_a^{-1}\mathbf{X}_{aK}\Omega_a^{-1}\mathbf{X}_{a1}' \\ \ldots & \ldots & \\ trV_a^{-1}\mathbf{X}_{a1}\Omega_a^{-1}\mathbf{X}_{aK}' & \ldots & trV_a^{-1}\mathbf{X}_{aK}\Omega_a^{-1}\mathbf{X}_{aK}', \end{pmatrix} \tag{26}$$

using the formula $trABCD = vec'D'(C' \otimes A)vecB$ *and* $vec'D = (vecD)'$ *denotes the row vectorization (see Magnus and Neudecker 1988). In the same way we find for the second* $(K \times 1)$ *cross-moment vector of the GLS estimate*

$$\mathbf{M}_{\tilde{X}\tilde{Y}} = \left(\mathbf{X}_a'(\Omega_a \otimes V_a)^{-1}vec(Y_a)\right) = \begin{pmatrix} trV_a^{-1}Y_a\Omega_a^{-1}\mathbf{X}_{a1}' \\ \ldots \\ trV_a^{-1}Y_a\Omega_a^{-1}\mathbf{X}_{aK}'. \end{pmatrix} \tag{27}$$

The estimated moment matrices $\hat{\mathbf{M}}$ *use the estimated covariance matrices* $\hat{\Omega}_d$ *and* $\hat{V}_d$.

3.2. Non-spatial Chow-Lin forecasts for SIF models

It was shown in Polasek et al. (2009) that the spatial Chow-Lin predictions have the form of a conditional mean for the disaggregated observations, given the aggregated model (the conditional density is denoted as $f(y)_{d|a}$), and yields the following "Chow-Lin formula".

Theorem 2. *The "Chow-Lin formula"*
The "Chow-Lin formula" for the missing disaggregated, given the observed aggregated observations is the conditional mean $\hat{y}_d$ *of the disaggregated observation in a joint system of observed and unobserved observation:*

$$\begin{aligned} \hat{\mathbf{y}}_d &= Plain + Gain * Residual \\ &= \hat{\mathbf{y}}_{d0} + \mathbf{AC}'(\mathbf{CAC}')^{-1}(\mathbf{y}_a - \hat{\mathbf{y}}_a), \end{aligned} \tag{28}$$

with $\hat{\mathbf{y}}_{d0} = fX_d\beta_d$ *and* $\hat{\mathbf{y}}_a = fCX_d\beta_d$ *is the fit from the ARF model.*

Proof 2. *The joint distribution of the aggregated and the disaggregated model is given by* $\mathbf{A} = Var(\mathbf{y}_d)$ *by*

$$\begin{pmatrix} \mathbf{y}_d \\ \mathbf{C}\mathbf{y}_d \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_d \\ \mu_a \end{pmatrix}, \sigma_d^2 \begin{pmatrix} \mathbf{A} & \mathbf{AC}' \\ \mathbf{CA} & \mathbf{CAC}' \end{pmatrix}\right]. \tag{29}$$

Since is this a partitioned normal distribution the conditional mean for the disaggregated data is given by

$$\mu_{d|a} = \mathbb{E}(\mathbf{y}_{d|a}) = \mu_d + \mathbf{AC}'(\mathbf{CAC}')^{-1}(\mathbf{y}_a - \hat{\mathbf{y}}_a) \tag{30}$$

while the conditional variance is

$$Var(\mathbf{y}_{d|a}) = \mathbf{A} - \mathbf{AC}'(\mathbf{CAC}')^{-1}\mathbf{CA} \tag{31}$$

In the nSIF model we have the following joint distribution between the y_d and y_a observations:

$$\begin{pmatrix} \mathbf{y}_d \\ \mathbf{y}_a \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_d \\ \mu_a \end{pmatrix}, \sigma_d^2 \begin{pmatrix} \Omega_d \otimes \mathbf{V}_d & (\Omega_d \otimes \mathbf{V}_d)\mathbf{C}' \\ ./. & \Omega_a \otimes \mathbf{V}_a \end{pmatrix}\right]. \tag{32}$$

This covariance between the aggregated and disaggregated residuals is

$$Cov(u_d, u_a) = E(u_d u_d' C') = \sigma_d^2(\Omega_d \otimes \mathbf{V}_d)C' = \sigma_d^2 \Sigma_d \mathbf{C}'. \tag{33}$$

In the SCL model the $\mathbf{C}$ matrix has a special diagonal structure $\mathbf{C} = diag(\mathbf{1}_{n_1}', ..., \mathbf{1}_{n_N}')$.

Now we find for the Chow-Lin formula in the nSIF model by assuming a homoskedastic error structure for the unknown disaggregated covariance matrix $\Sigma_d = \mathbf{I}_{nn}$, which is matrix A in formula (29).

The non-spatial Chow-Lin forecasts in the nSIF model are also given by the Chow-Lin formula (28) and the covariance matrices in the heteroskedastic case have to be estimated. Another way of avoiding the assumption is to parameterize the covariance matrices by a distance correlation function. Let $\mathbf{W} : n \times n$ be a known positive (non-negative and symmetric) distance matrix with zeros in the main diagonal, then for $0 \le \rho < 1$ we define the correlation matrix $\mathbf{S} = \rho^{-\mathbf{D}}$. $\mathbf{S}$ has 1's in the main diagonal and all other entries are between 0 and 1.
We can parameterize the covariance matrix now e.g. as $\mathbf{V}_d = \mathbf{D}_\sigma(\mathbf{I}_n + \rho^{-\mathbf{W}_d})\mathbf{D}_\sigma$ where $\mathbf{D}_\sigma = diag(\sigma_1, ..., \sigma_n)$ is a diagonal matrix of n standard deviations and the matrix exponent in $\rho^{-\mathbf{W}_d}$ is understood to be point-wise, yielding a $n \times n$ matrix. All together we would have to estimate $N + 1$ parameters from the aggregate model and then make the assumption that the disaggregate covariance matrix can be 'extrapolated' by $V_d = \mathbf{D}_\sigma(I_n + \rho_a^{-W_d})\mathbf{D}_\sigma$. Since the disaggregate standard deviations of the D_σ vector are unknown we have to make the Chow-Lin type of dilution assumption: The disaggregate standard deviations of the subunits are equal to the aggregate standard deviations in this aggregation unit. Thus, in analogy a similar Gain*Residual formula can be used:

$$\sigma_d = \mathbf{D}_{d,\sigma}\mathbf{1}_n = \mathbf{C}\ \mathbf{D}_n^{-1}\mathbf{C}\mathbf{D}_{a,\sigma}1_N.$$

Here σ_d is a $n \times 1$ vector of disaggregate standard deviations and $\sigma_a = \mathbf{D}_{a,\sigma}\mathbf{1}_N$ is an $N \times 1$ aggregate vector of standard deviations.

3.3. Feasible generalized least squares (FGLS) estimation in the nSIF model

Consider the aggregated homoskedastic nSIF model in panel form:

$$\mathbf{Y}_a = \sum_{i}^{K} \mathbf{X}_{ai}\beta_{di} + U_a, \quad U_a \sim \mathcal{N}_{N\times N}[\mathbf{0}, \sigma_d^2 I_{NN}], \tag{34}$$

where the aggregates are given by

$$Y_a = \mathbf{C}_0 Y_d \mathbf{C}_0' \quad \text{and} \quad \mathbf{X}_{ai} = \mathbf{C}_0 \mathbf{X}_{di} \mathbf{C}_0', \quad i = 1, \ldots, K,$$

from the disaggregated panels and with $I_{NN} = I_N \otimes I_N$. Next, we need the aggregated model equation to estimate β_d. Note that by aggregation we get a heteroskedastic model with the variance matrix

$$\Sigma_a = Var(\mathbf{C}vec(U_a)) = \mathbf{C}Var(u_a)\mathbf{C}' = \sigma_d^2(\mathbf{C}_0\mathbf{C}_0' \otimes \mathbf{C}_0\mathbf{C}_0') = \sigma_d^2 \mathbf{D}_{nn},$$

where $\mathbf{D}_{nn} = \mathbf{D}_n \otimes \mathbf{D}_n$ is diagonal since $\mathbf{D}_n = \mathbf{C}_0\mathbf{C}_0' = diag(d_1, \ldots, d_n)$ is a diagonal matrix of positive numbers.

Therefore the $K \times K$ regressor moment matrix $\mathbf{M}_{XX}$ is given via Theorem (1) where the elements are computed as numbers from trace operations:

$$\mathbf{M}_{XX} = \mathbf{X}_a'\Sigma_a^{-1}\mathbf{X}_a = \left[tr\mathbf{X}_{ai}'\mathbf{D}_n^{-1}\mathbf{X}_{aj}\mathbf{D}_n^{-1}\right]_{i,j=1,\ldots,K}, \tag{35}$$

and the cross-moment vector is via (27)

$$\mathbf{M}_{XY} = \mathbf{X}_a'\Sigma_a^{-1}y_a = \left[tr\mathbf{X}_{ai}'\mathbf{D}_n^{-1}Y_a\mathbf{D}_n^{-1}\right]_{i=1,\ldots,K}. \tag{36}$$

Because of the diagonal structure we can call this estimator the weighted or WLS estimator and the nSIF least squares estimator can be computed as in (24). We summarize this result in the next theorem.

Theorem 3 (WLS in the nSIF model). *The LS estimate in the nSIF model (34) is given by*

$$\beta_d^{nSIF} = \mathbf{M}_{XX}^{-1}\mathbf{M}_{XY}$$

with the moments given in (35) and (36).

Proof 3. *Follows from the above.*

The residuals of the WLS estimates in the nSIF model are

$$\hat{\mathbf{U}}_a = \mathbf{Y}_a - \sum_{i=1}^{K} \mathbf{X}_a \hat{\beta}_{ai}^{WLS},$$

and this estimate can be also used to make Chow-Lin predictions for the nSIF model as in (37):

$$\hat{\mathbf{y}}_d^{nSIF} = \hat{\mathbf{y}}_{d0} + \mathbf{C}'\mathbf{D}_n^{-1}\mathbf{C}\mathbf{u}_a^{nSIF},$$

with $\mathbf{u}_a^{nSIF} = \mathbf{y}_a - \tilde{\mathbf{X}}_a\hat{\boldsymbol{\beta}}_d^{nSIF}$ and $\tilde{\mathbf{X}}_a : NN \times K$ is the aggregated regressor matrix.

Note that it is possible to use $\hat{\mathbf{y}}_d^{nSIF}$ to construct a full covariance matrix for the disaggregated model.

4. The general origin-destination spatial internal flow (OD-SIF) model

In this section we show how the LS estimation works in simple and complicated spatial SIF models. We start with the non-spatial OD-SIF regression model for the aggregated observations:

$$\mathbf{Y}_d = \sum_{i=1}^{K} \mathbf{X}_{di}\beta_{di} + \mathbf{U}_d, \quad U_d \sim \mathcal{N}_{n\times n}[\mathbf{0}, \sigma_d^2\Omega_d \otimes \mathbf{V}_d], \tag{37}$$

where K is the number of regressors and $\mathcal{N}_{n\times n}$ denotes the matrix normal distribution, and there are K disaggregated regressor panels $\mathbf{X}_{di} : n \times n, i = 1, ..., K$. In contrast, the OD-SIF model in matrix form for aggregated data has the form

$$\mathbf{Y}_a = \sum_{i=1}^{K} \mathbf{X}_{ai}\beta_{ai} + U_a, \quad U_a \sim \mathcal{N}_{n\times n}[\mathbf{0}, \sigma_a^2\Omega_{agg} \otimes V_{agg}], \tag{38}$$

where $\mathbf{Y}_a = \mathbf{C}\mathbf{Y}_d\mathbf{C}' : N \times N$.

Next we extend the non-spatial model (37) with spatial lags. A general 3-component spatial lag polynomial for OD regressions can be defined by

$$\begin{aligned}\mathbf{R}_\rho &= I_{nn} - \rho_1(W_1 \otimes I_n) - \rho_2(I_n \otimes W_2) + \rho_3(W_1 \otimes W_2) = \\ &= \tilde{\mathbf{R}}_{\rho1} + \tilde{\mathbf{R}}_{\rho2} - \tilde{\mathbf{R}}_{\rho3},\end{aligned} \tag{39}$$

with the following 3 components

$$\begin{aligned}\tilde{\mathbf{R}}_{\rho1} &= I_{nn} - \rho_1(W_1 \otimes I_n) = R_1 \otimes I_n \\ \tilde{\mathbf{R}}_{\rho2} &= I_{nn} - \rho_2(I_n \otimes W_2) = I_n \otimes R_2 \\ \tilde{\mathbf{R}}_{\rho3} &= I_{nn} - \rho_3(W_1 \otimes W_2) = \tilde{R}_1 \otimes \tilde{R}_2,\end{aligned} \tag{40}$$

and the spread matrices are defined for each ρ-component:

$$\mathbf{R}_i = I_n - \rho_i W_i, \quad i = 1, 2 \quad \text{and} \quad \tilde{R}_i = I_n - \sqrt{\rho_3} W_i, i = 1, 2.$$

The OD-SAR model is an OD-SIF model that uses 3 spatial neighborhood components as spatial lags

$$\mathbf{Y}_d = \mathbf{W}_\rho \mathbf{Y}_d + \sum_{i=1}^{K} \mathbf{X}_{di}\beta_{di} + \mathbf{U}_d, \quad \mathbf{U}_d \sim \mathcal{N}_{n\times n}[\mathbf{0}, \sigma_d^2 \Omega_d \otimes \mathbf{V}_d] \tag{41}$$

with the OD-polynomial

$$\mathbf{W}_\rho = \rho(\mathbf{W}_1, \mathbf{W}_2) = \rho_1(\mathbf{W}_1 \otimes I_n) - \rho_2(I_n \otimes \mathbf{W}_2) + \rho_3(\mathbf{W}_1 \otimes \mathbf{W}_2). \tag{42}$$

The feasible GLS estimator for β_d using the aggregate model is given by

$$\hat{\beta}_d^{FGLS} \quad = \quad (\mathbf{X}_d'\mathbf{C}'(\mathbf{C}\hat{V}_{a\rho}\mathbf{C}')^{-1}CX_d)^{-1}\mathbf{X}_d'\mathbf{C}'(\mathbf{C}\hat{V}_{a\rho}\mathbf{C}')^{-1}y_a, \tag{43}$$

with the estimated covariance matrix from the aggregated reduced form of the SAR model

$$\hat{V}_{a\rho} = \mathbf{C}\hat{\mathbf{R}}_\rho^{-1}(\hat{\Omega} \otimes \hat{V})\hat{\mathbf{R}}_\rho'^{-1}\mathbf{C}',$$

(see Polasek et al., 2009), and where we have replaced the unknown parameters in (13) by their estimates. Estimation of the ρ_i coefficients can be done numerically over a 3-dimensional grid, but it is computationally intensive. The Chow-Lin formula (i.e. the BLUE prediction of the missing disaggregated values) for the flow SAR model is now given for the disaggregated model

$$\begin{aligned} \hat{y}_d &= \mathbf{R}_\rho^{-1}X_d\hat{\beta}_{GLS} + \hat{V}_{a\rho}\mathbf{C}'(\mathbf{C}\hat{V}_{a\rho}\mathbf{C}')^{-1}(y_a - \mathbf{C}\mathbf{R}_\rho^{-1}X_d\hat{\beta}_{GLS}) = \\ &= \hat{y}_0 + \mathbf{G}\hat{u}_a = \\ &= Plain + Gain * Residual, \end{aligned} \tag{44}$$

where the variables are defined in the same way as in the non-spatial Chow-Lin model (4). The spatial improvement of the Goldberger (1962) 'gain projection matrix' is now

$$\mathbf{G} = \hat{V}_{a\rho}\mathbf{C}'(\mathbf{C}\hat{V}_{a\rho}\mathbf{C}')^{-1}, \tag{45}$$

and distributes the estimated aggregate residuals $\hat{u}_a = y_a - \mathbf{C}\mathbf{R}_{\hat{\rho}}^{-1}X_d\hat{\beta}_{GLS}$ across the spatial naive prediction $\hat{y}_0 = \mathbf{R}_{\hat{\rho}}^{-1}X_d\hat{\beta}_{GLS}$.

Instead of assuming the whole spatial lag polynomial (39) we could find an easier way and estimate the components individually. In the next subsections we will discuss the general case and the special cases for estimation.

4.1. The origin spatial internal flow (oSIF) model

In this section we consider the uni-lateral "origin-only" spatial internal flow (oSIF) model as in the special form of the general SIF model (6). Thus the oSIF model uses only the origin component of the lag polynomial to define the spatial origin lag variable:

$$\rho(\mathbf{W}_1, \mathbf{W}_2) y_d = \rho_1 (\mathbf{W}_1 \otimes I_n) y_d = \tilde{\mathbf{W}}_1 y_d = vec(Y_d \mathbf{W}_1'), \tag{46}$$

with $\tilde{\mathbf{W}}_1 = \mathbf{W}_1 \otimes I_n$. The heteroskedastic SAR-oSIF model is defined as

$$y_d = \rho_1 \tilde{\mathbf{W}}_1 y_d + \mathbf{X}_d \beta_d + \mathbf{u}_d, \quad \mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2 \Omega_d \otimes \mathbf{V}_d], \tag{47}$$

or in matrix notation we can write using (46)

$$Y_d = \rho_1 Y_d \mathbf{W}_1' + \sum_i \mathbf{X}_{di} \beta_{di} + U_d, \quad U_d \sim \mathcal{N}_{n \times n}[\mathbf{0}, \sigma_d^2 \mathbf{\Sigma}_d], \tag{48}$$

where $\mathbf{U}_d : n \times n$ is the residual matrix of the flow model and the full (heteroskedastic) covariance matrix of the flow model is

$$\Sigma_d = \Omega_d \otimes \mathbf{V}_d, \tag{49}$$

while for the homoskedastic covariance matrix of the disaggregated flow model we assume

$$Var(\mathbf{u}_d) = \sigma_d^2 I_n \otimes I_n. \tag{50}$$

The reduced form of the oSIF model is

$$y_d \sim \mathcal{N}[\tilde{\mathbf{R}}^{-1} \mathbf{X}_d \beta_d, \Sigma_{d1} = \sigma_d^2 \Omega_1 \otimes \mathbf{V}_d], \tag{51}$$

with $\tilde{\mathbf{R}} = \mathbf{R}_1 \otimes I_n$ and $\mathbf{R}_1 = I_n - \rho_1 W_1$ and

$$\Omega_1 = \mathbf{R}_1^{-1} \Omega_d \mathbf{R}_1^{'-1}, \tag{52}$$

because

$$\Sigma_{d1} = (\mathbf{R}_1^{-1} \otimes I_n)(\Omega_d \otimes \mathbf{V}_d)(\mathbf{R}_1^{'-1} \otimes I_n) = \mathbf{R}_1^{-1} \Omega \mathbf{R}_1^{'-1} \otimes \mathbf{V}_d = \Omega_1 \otimes V_d. \tag{53}$$

In matrix form the reduced form of the oSIF model is

$$Y_d \sim \mathcal{N}_{n \times n} \left[\sum_{i=1}^{K} \mathbf{X}_{di} \mathbf{R}_1^{'-1} \beta_i, \sigma_d^2 \Sigma_{d1} \right]. \tag{54}$$

The aggregated reduced form of the oSIF model in (47) is given by multiplying the reduced form by the aggregation matrix $\mathbf{C}$. Thus, the ARF-oSIF model has the following form:

$$\mathbf{C}\mathbf{y}_d = \mathbf{C}\mathbf{R}^{-1} X_d \beta_d + \mathbf{C}\mathbf{R}^{-1} u_d, \quad \mathbf{C}\mathbf{R}^{-1} u_d \sim \mathcal{N}[0, \Sigma_{d2}], \tag{55}$$

where the spread matrix $\mathbf{R}_1$ for oSIF flows is given in (40) and with

$$\Sigma_{d2} = \mathbf{C}(\Omega_1 \otimes V_d)\mathbf{C}' = \mathbf{C}_0 \Omega_1 \mathbf{C}_0' \otimes \mathbf{C}_0 V_d \mathbf{C}_0' = \Omega_2 \otimes V_2. \tag{56}$$

The estimated covariance matrix replaces the unknown parameters by ML estimates

$$\hat{\Sigma}_{d2} = \hat{\Omega}_2 \otimes \hat{V}_2. \tag{57}$$

In matrix form the ARF model can be written with $\mathbf{Y}_a = \mathbf{C}C_0 Y_d \mathbf{C}_0'$ as

$$\mathbf{Y}_a \sim \mathcal{N}_{n\times n}[\sum_{i=1}^{K} \mathbf{C}_0 \mathbf{X}_{di} \mathbf{R}_1^{'-1} \mathbf{C}_0' \beta_{di}, \sigma_d^2 \Sigma_{d2}]. \tag{58}$$

Next we turn to the problem of how to estimate the parameters in an oSIF model $\theta_d = (\beta_d, \sigma_d^2, \rho_d)$ by the existing SAR programs (e.g. in the packages R or MATLAB). This leads to the following FGLS procedure.

Procedure 1 ($\hat{\beta}_d$: Feasible GLS in the SAR-oSIF model). *The feasible GLS estimation of the aggregated reduced form (ARF) model (55) is given by*

$$\hat{\beta}_d^{GLS} = (X_{d2}' \hat{\Sigma}_{d2}^{-1} X_{d2})^{-1} X_{d2}' \hat{\Sigma}_{d2}^{-1} y_a, \tag{59}$$

with $X_{d2} = \mathbf{C}\mathbf{R}^{-1} X_d$ *and the estimated covariance matrix is*

$$\hat{\Sigma}_{d2} = \mathbf{C}_0 \hat{\mathbf{R}}_1^{-1} \hat{\Omega}_d \hat{\mathbf{R}}_1^{'-1} \mathbf{C}_0' \otimes \mathbf{C}_0 \hat{V} \mathbf{C}_0'. \tag{60}$$

The feasible GLS (FGLS) procedure can be set up in the following way:

- *Estimate* $\hat{\beta}_d^{WLS}$ *by the homoskedastic nSIF flow model with* $\Sigma_d = \mathbf{I}_{nn}$ *as in (50).*
- *Compute* $\hat{\Sigma}_{d2}$ *using the residuals of the homoskedastic nSIF flow model:* $\hat{\Omega}_d = \mathbf{U}_a' \mathbf{U}_a / N$
- *Make a Cholesky decomposition of* $\hat{\Sigma}_{d2} = S'S \otimes L'L$.
- *Compute the transformed regressors*
  $Y^* = L^{'-1} Y_a S^{-1}$ *and* $X_i^* = L^{'-1} X_i S^{-1}, i = 1, \ldots, K$.
- *Estimate* $\hat{\beta}_d^{FGLS}$ *by applying a SAR model with the transformed regressors* Y^* *and* X_i^*.

The rationale behind this procedure is: Insert the ARF model (55) into the GLS estimation formula and approximate the unknown correlation structure in a step-wise estimation procedure.

Note 1 (Feasible GLS for the homoskedastic ARF model). *The GLS estimation formula is simplified if we assume homoskedastic covariance matrices for the oSIF model as in (50).*

Then $\Sigma_{a\rho}$ *can be simplified to the homoskedastic case by assuming*

$$\Sigma_{a\rho} = C \tilde{\mathbf{R}}_1^{-1} (\sigma^2 I_{nn}) \tilde{\mathbf{R}}_1^{'-1} C' = \sigma_d^2 C (\tilde{\mathbf{R}}_1' \tilde{\mathbf{R}}_1)^{-1} C' = \sigma_d^2 \Sigma_{a1} \otimes D_n, \tag{61}$$

because

$$\hat{\Sigma}_{a\rho} = \sigma_d^2 \mathbf{C}((\mathbf{R}_1' \mathbf{R}_1)^{-1} \otimes I_n) \mathbf{C}' = \sigma_d^2 \mathbf{C}_0 (\mathbf{R}_1' \mathbf{R}_1)^{-1} \mathbf{C}_0' \otimes \mathbf{C}_0 \mathbf{C}_0',$$

since $\mathbf{D}_n = \mathbf{C}_0\mathbf{C}_0'$ *and with* $\hat{\Sigma}_{a1}$ *the aggregated covariance matrix given by*

$$\hat{\Sigma}_{a1} = \mathbf{C}\hat{\Sigma}_{d1}\mathbf{C}' = \hat{\sigma}_d^2\mathbf{C}_0\hat{\Omega}_1\mathbf{C}_0' \otimes \mathbf{C}_0\hat{V}\mathbf{C}_0'. \tag{62}$$

The ML estimates of the covariance matrices are

$$\hat{\Omega} = \hat{U}_a\hat{U}_a'/N \quad \textit{and} \quad \hat{V} = \hat{U}_a'\hat{U}_a/n, \tag{63}$$

where $\hat{U}_a$ *is the residual matrix of the homoskedastic model:* $\hat{U}_a = Y_a - \hat{Y}_0$ *where* $\hat{Y}_0$ *is the plain OLS prediction, assuming a homoskedastic error structure as in (50).*

In similar way we find the FGLS for the dSIF model.

Theorem 4 ($\hat{\beta}_d$: Feasible GLS in the dSIF model). *The GLS estimator of the aggregated reduced form (ARF) model (55) is given by*

$$\hat{\beta}_d^{GLS} = (X_{d2}'\hat{\Sigma}_{d2}^{-1}\mathbf{X}_{d2})^{-1}\mathbf{X}_{d2}'\hat{\Sigma}_{d2}^{-1}y_a, \tag{64}$$

with $\mathbf{X}_{d2} = \mathbf{C}\mathbf{R}^{-1}\mathbf{X}_d$ *with* $\mathbf{R} = \mathbf{R}_2 \otimes \mathbf{I}_N$ *and the estimated covariance matrix*

$$\hat{\Sigma}_{d2} = \mathbf{C}_0\hat{\Omega}\mathbf{C}_0' \otimes \mathbf{C}_0\hat{R}_2^{-1}\hat{V}\hat{R}_2^{'-1}\mathbf{C}_0'. \tag{65}$$

Proof 4. *Follows the proof of the oSIF model.*

This suggests the feasible GLS (FGLS) procedure in the same way as in Procedure 1, but now the second step replaced by:

- Estimate $\hat{\mathbf{R}}_2 = \mathbf{I}_N - \hat{\rho}\mathbf{W}_2$ and construct $\hat{\Sigma}_{d2}$ for the dSIF model.

4.2. Chow-Lin predictions in the SAR-SIF model

The FGLS results of the previous section can be used for the Chow-Lin prediction in the disaggregate model:

$$\hat{y}_d = \mathbf{R}^{-1}\mathbf{X}_d\hat{\beta}_{GLS} + \hat{\Sigma}_{d1}\mathbf{C}'(\mathbf{C}\hat{\Sigma}_{d1}\mathbf{C}')^{-1}(\mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}X_d\hat{\beta}_{GLS}). \tag{66}$$

The plain point forecasts are computed with the GLS estimate $\hat{y}_0 = (\mathbf{R}_1^{-1} \otimes \mathbf{I}_n)\mathbf{X}_d\hat{\beta}_{GLS}$ or

$$vec\hat{Y}_0 = \sum_{j=1}^{n} vec(\mathbf{X}_{dj}\mathbf{R}_1^{'-1})\hat{\beta}_{GLS,j}$$

and the 'Goldberger gain matrix' stems from the ARF in (65) and is derived via the covariance matrices that are used in the Chow-Lin approach in the same way as in (33). The improvement term is

$$Gain * Residual \quad = \quad (\hat{\Omega} \otimes \mathbf{C}_0'\hat{R}_2^{-1}\hat{V}\hat{R}_2^{'-1}\mathbf{C}_0')\hat{\Sigma}_{d2}^{-1}\hat{u}_a^{GLS}, \tag{67}$$

with $\hat{\Sigma}_{d2} = \mathbf{C}_0\hat{\Omega}\mathbf{C}_0' \otimes \mathbf{C}_0\hat{R}_2^{-1}\hat{V}_d\hat{R}_2^{'-1}\mathbf{C}_0'$ and $\hat{u}_a^{GLS} = \mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}\mathbf{X}_d\hat{\beta}_{GLS}$.

Therefore we can summarize the least squares prediction in the oSIF model for O/D-matrices in the following way using the spatial CL (SCL) approach of Polasek et al. (2009):

Procedure 2 (Chow-Lin prediction in the oSIF model). *We consider the SIF model (6)*

1. *Vectorize the aggregate* $\mathbf{Y}_a$ *and* $\mathbf{X}_a$ *matrices and run the ordinary SAR-SCL program.*
2. *Compute the simple ('plain') aggregate residual* $\hat{\mathbf{U}}_0 = \mathbf{Y}_0 - \hat{\mathbf{Y}}_0$ *and the covariance matrices* $\hat{\Sigma}$ *and* $\hat{\mathbf{V}}$ *in (63).*
3. *Estimate* $\hat{\beta}_{d,FGLS}$ *as in (43).*
4. *Compute the Chow-Lin forecasts (66) with the known* X_d *matrices.*

Note: The Chow-Lin prediction in the dSIF model follows parallel steps as in the oSIF model.

4.3. Estimation with structural zeros in trade flow models

If we estimate flow models with trade, the trade within a cell is recorded by a 0 and so the y-observation at this location is zero (structural zero). For the estimation (to avoid biases) we have to ignore these values and they are deleted from the model (e.g. SAR) estimation. We outline the procedure by an example with a 10×10 trade flow matrix. To avoid biases in the estimation we need to eliminate the observation with a structural zero in the vectorized regression.

Procedure 3 (Estimation with structural zeros).

1. *Vectorize all variables and eliminate every 10th observation, giving 90 non-zero observations.*
2. *Eliminate the corresponding rows in the* Σ_ρ *matrix.*
3. *Estimate* β *and* ρ *from the non-zero system and get the residual vector* $\mathbf{u}$*.*
4. *Construct the residual matrix* $\mathbf{U}$ *by inserting into the main diagonal 'NA's.*
5. *Estimate the within and between covariance matrices by a 'NA' procedure (skipping over non fully observed pairs).*
6. *Make a Cholesky transformation and transform the original variables.*
7. *Eliminate again all observations that correspond to the structural zeros and estimate the remaining system by a homoskedastic procedure.*

4.4. Uni-lateral spatial GLS estimation in the SAR-SIF model

In this section we will explore the estimation of the uni-lateral spatial SAR-SIF models that will only consider the neighborhood relationships at the origin or destination separately. The reason for this a simplification in the estimation formulas involved.[1]

First, we consider the SIF model with an additional simple spatial origin lag of the form $\mathbf{W}_2\mathbf{Y}_a$ or $\mathbf{Y}_a\mathbf{W}_1'$. No simple matrix expression for the joint origin-destination lag $\mathbf{W}_2\mathbf{Y}_a\mathbf{W}_1'$ is possible. For such a model we can only estimate the SIF model in vectorized form, and thus the size of the matrices is dependent on the storage capabilities of the computing environment.

The disaggregate SAR-oSIF model with $\mathbf{W} = \mathbf{W}_1$ is

$$\mathbf{Y}_d = \rho \mathbf{W}\mathbf{Y}_d + \sum_{i=1}^{K} \mathbf{X}_{di}\beta_{di} + \mathbf{U}_d, \quad \mathbf{U}_d \sim \mathcal{N}[\mathbf{0}, \Omega_d \otimes \mathbf{V}_d], \tag{68}$$

where $\mathbf{Y}_d : N \times N$ and β_{di} is the i-th element of the regression vector $\beta_d : K \times 1$. The vectorized form of the model is, with $\mathbf{y}_d = vec\mathbf{Y}_d$, $\mathbf{u}_d = vec\mathbf{U}_d$ and $\mathbf{x}_{di} = vec\mathbf{X}_{di}, i = 1, \ldots, K$,

$$\mathbf{y}_d = \rho(I_n \otimes \mathbf{W})\mathbf{y}_d + \sum_{i=1}^{K} \mathbf{x}_{di}\beta_{di} + \mathbf{u}_d, \quad \mathbf{u}_d \sim \mathcal{N}[\mathbf{0}, \Omega_d \otimes \mathbf{V}_d]. \tag{69}$$

The ARF of the oSIF model uses the (origin-lateral) spread matrix $R = I_{nn} - \rho(I_n \otimes \mathbf{W}) = I_n \otimes R_1$ and is given by

$$\mathbf{Y}_a \sim \mathcal{N}[\sum_{i=1}^{K} \mathbf{R}_1^{-1}\mathbf{X}_a\beta_{di}, \mathbf{C}_o = \mathbf{C}(\Omega \otimes \mathbf{V}_r)\mathbf{C}'], \tag{70}$$

with the aggregated observations $\mathbf{Y}_a = \mathbf{C}\mathbf{Y}_d\mathbf{C}' : n \times n$, $\mathbf{X}_r = \mathbf{C}R_1^{-1}\mathbf{X}_d C' : n \times n, \mathbf{V}_r = (R_1' V^{-1} R_1)^{-1}$ and $R_1 = I_n - \rho W$. And we get for the aggregated covariance matrix $\mathbf{V}_o$

$$\mathbf{V}_o = (\mathbf{C}_0\Omega\mathbf{C}_0') \otimes (\mathbf{C}_0\mathbf{V}_r\mathbf{C}_0') = \Omega_1 \otimes V_1. \tag{71}$$

Since the ARF of the oSIF model has the same structural form as the non-spatial SIF model, we can use the GLS estimation of Theorem (1) with the GLS estimate given by $\beta_{GLS} = \mathbf{M}_{XX}^{-1}\mathbf{M}_{XY}$.

This implies the element-wise construction of the moment matrix (26), computed as numbers from trace operations:

$$M_{XX} = X_a'\mathbf{V}_o^{-1}X_a = \left(trX_{ai}'\mathbf{V}_1^{-1}X_{aj}\Omega_1^{-1}\right)_{i,j=1,\ldots,K}, \tag{72}$$

with the variance matrix $\mathbf{V}_o$ from the ARF form (70) and the cross-moment vector as in (27) by

$$M_{XY} = X_a'\mathbf{V}_o^{-1}y_a = \left[trX_{ai}'\mathbf{V}_1^{-1}Y_a\Omega_1^{-1}\right]_{i=1,\ldots,K}. \tag{73}$$

[1] A bilateral spatial SAR-SIF model is defined by a spatial polynomial that takes into account the spatial neighborhood relationships at the origin and destination simultaneously.

Similarly, as in the oSIF model, the aggregated reduced form (ARF) of the SAR-dSIF model is given by

$$Y_a \sim \mathcal{N}\left[\sum_{i=1}^{K} X_a \beta_{di}, \Omega_r \otimes \mathbf{V}\right] \tag{74}$$

with $\Omega_r = (R_1' \Omega^{-1} R_1)^{-1}$ and $R_1 = I_n - \rho_d W$ as before.

Again the ARF of the dSIF model has the same structural from as the non-spatial SIF model, we can use the GLS estimation of Theorem (1) with the GLS estimate given by the moment matrix

$$\mathbf{M}_{XX} = X_a' \mathbf{V}_d^{-1} X_a = \left[tr X_{ai}' \mathbf{V}_1^{-1} X_{aj} \Omega_1^{-1}\right]_{i,j=1,\ldots,K}, \tag{75}$$

with the $\mathbf{V}_d = \mathbf{C}(\Omega_r \otimes \mathbf{V})\mathbf{C}'$ from the ARF form (74) and the cross-moment vector as in (27) by

$$\mathbf{M}_{XY} = X_a' \mathbf{V}_d^{-1} y_a = \left[tr X_{ai}' \mathbf{V}_1^{-1} Y_a \Omega_1^{-1}\right]_{i=1,\ldots,K}. \tag{76}$$

4.5. Feasible GLS estimation in unilateral models

The feasible GLS estimator of β_a is given by

$$\beta_a^{FGLS} = (X_a' \hat{\mathbf{V}}_{a1}^{-1} X_a)^{-1} X_a' \hat{\mathbf{V}}_{a1}^{-1} y_a$$

with

$$\mathbf{V}_{a1} = \hat{\Omega}_a \otimes \hat{\mathbf{V}}_1, \quad \textit{with} \quad \hat{\mathbf{V}}_1 = (\mathbf{R}_1' \hat{\mathbf{V}}_a^{-1} \mathbf{R}_1)^{-1}. \tag{77}$$

For the estimation of the residual variance-covariance matrices we define the matrices $\hat{\Omega}_a = \hat{U}_a' \hat{U}_a / N$ and $\hat{\mathbf{V}}_a = \hat{U}_a \hat{U}_a' / N$ as before but now with the non-spatially estimated residuals $\hat{U}_a = Y_a - \sum_i^K X_a \hat{\beta}_{di}$. In case of the dSIF model we have

$$\mathbf{V}_{a1} = \hat{\Omega}_1 \otimes \mathbf{V}. \tag{78}$$

For bilateral models the FGLS estimator is with $y_a = vecY_a$

$$\mathbf{M}_{\tilde{X}\tilde{Y}} = \left[X_a' (\Omega_a \otimes \mathbf{V}_a)^{-1} vec(W_1 Y_a W_2')\right] = \begin{pmatrix} tr \mathbf{V}_a^{-1} W_1 Y_a W_2' \Omega_a^{-1} X_{a1}' \\ \ldots \\ tr \mathbf{V}_a^{-1} W_1 Y_a W_2' \Omega_a^{-1} X_{aK}' \end{pmatrix}.$$

For the estimation of ρ_d a grid search for ρ_d is possible: The minimum of the spatial ρ is found by minimizing over a grid of rho values in the interval $(-1, 1)$.

4.6. Chow-Lin prediction in the SIF model

The Chow-Lin forecasting has to be done by using the usual Goldberger (1962) formula: $\hat{y}_d = X_d\beta_d^{GLS} + G\hat{u}_a$ where the term $G\hat{u}_a$ is an improvement of the estimated error term $\hat{u}_a = (y_a - X_{a\rho}\beta_d^{GLS})$ using the "'Goldberger gain"' matrix $G = \mathbf{V}_{a\rho}^{-1}\mathbf{C}'(\mathbf{C}\mathbf{V}_{a\rho}^{-1}\mathbf{C}')^{-1}$.

The point forecasts can be calculated as

$$\hat{y}_d = \mathbf{X}\beta_d^{GLS} + \mathbf{G}\hat{\mathbf{u}}_a, \tag{79}$$

and the matrix $\mathbf{Y}$ is obtained by de-vectorizing: $\hat{\mathbf{y}} = vec\hat{\mathbf{Y}}$.

5. Unilateral spatial lags in the Bayesian SAR-SIF model

For the Bayesian treatment of the oSIF or dSIF model (68) we have to assume a prior distribution and then we just adopt the heteroskedastic MCMC algorithm.

The Bayesian estimation follows the same line as in the general SIF model and can be summarized by the MCMC procedure similar as in Theorem 3. We consider the model with the prior distribution for $\Theta = (\beta, \rho, \phi, \sigma^2, \Omega)$

$$p(\Theta) = \prod_{i=1}^{n} \mathcal{N}[\boldsymbol{\beta}_i \mid \mathbf{b}_i^*, \mathbf{H}_*]\ \mathcal{W}_N[\Omega^{-1} \mid (\nu_*\boldsymbol{\Omega}_*)^{-1}, \nu_*]\ \mathcal{U}_{-1,1}(\boldsymbol{\rho})\ \mathcal{U}_{-1,1}(\boldsymbol{\phi}), \tag{80}$$

where $\mathcal{U}_{-1,1}(\boldsymbol{\rho})$ and $\mathcal{U}_{-1,1}(\boldsymbol{\phi})$ stands for a uniform distributions in the interval $(-1, 1)$ and $\mathcal{W}_N$ for a Wishart distribution of dimension N. Note that simplified formulas emerge, if we assume a Zellner type "g-prior" of for the betas, centered at zero:

$$p(\boldsymbol{\beta}_d) \quad = \quad \mathcal{N}[\boldsymbol{\beta}_d \mid \mathbf{0}, \mathbf{H}_* = g\mathbf{I}_K], \tag{81}$$

with the scalar g being large (e.g. $g = 10^3$ or 10^6).

The likelihood function is given by the aggregated reduced form of the spatial internal flow (ARF-SIF) model for $\mathbf{C}\mathbf{y}_d = \mathbf{y}_a$

$$\mathbf{y}_a \sim \mathcal{N}[C\tilde{\mathbf{R}}^{-1}\mathbf{X}_d\beta_d, C\tilde{\mathbf{R}}^{-1}(\Omega \otimes \sigma^2\mathbf{V})\tilde{\mathbf{R}}'^{-1}C']. \tag{82}$$

Then the joint posterior distribution of the disaggregated model parameter Θ_d can be simulated numerically by MCMC.

Theorem 5 (MCMC in the internal flow model). *The MCMC in the spatial internal flow (SIF) model for the disaggregated model parameter is implemented as follows (for simplicity the d-index is omitted from the parameters)*

1. *Draw $\boldsymbol{\beta}$ from $\mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_{**}, \mathbf{H}_{**}]$*
2. *Draw ρ by a Metropolis step: $\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau_1^2]$*

3. *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s^2_{**}n_{**}/2, n_{**}/2]$
4. *Draw* $\mathbf{\Omega}^{-1}$ *from* $p(\mathbf{\Omega}^{-1} \mid \mathbf{Y}, \Theta^c) = \mathcal{W}_N[\mathbf{\Omega}^{-1} \mid (\nu_{**}\mathbf{\Omega}_{**})^{-1}, \nu_{**}]$
5. *Repeat until convergence.*

Proof 5 (Proof of Theorem 5).

(a) *The fcd for the beta regression coefficients is*

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbf{y}, \theta^c) &= \mathcal{N}[\boldsymbol{\beta} \mid \mathbf{b}_*, \mathbf{H}_*] \cdot \mathcal{N}[\mathbf{Cy} \mid \mathbf{CR}^{-1}\mathbf{X}\beta, \sigma^2\mathbf{V}_{a\rho}] \\ &= \mathcal{N}\left[\boldsymbol{\beta} \mid \mathbf{b}_{**}, \mathbf{H}_{**}\right], \end{aligned} \tag{83}$$

with the parameters

$$\begin{aligned} \mathbf{H}^{-1}_{**} &= \mathbf{H}^{-1}_* + \sigma^{-2}\mathbf{X}^\top R'^{-1}C'\mathbf{V}^{-1}_{a\rho}CR^{-1}\mathbf{X}, \\ \mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}^{-1}_*\mathbf{b}_* + \sigma^{-2}\mathbf{X}'\mathbf{R}'^{-1}\mathbf{C}'\mathbf{V}^{-1}_{a\rho}\mathbf{Cy}], \end{aligned} \tag{84}$$

using (78). These formulas can be written as

$$\begin{aligned} \mathbf{H}^{-1}_{**} &= \mathbf{H}^{-1}_* + \sigma^{-2}\mathbf{M}_{\tilde{X}R\tilde{X}}, \\ \mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}^{-1}_*\mathbf{b}_* + \sigma^{-2}\mathbf{M}_{\tilde{X}R\tilde{Y}}], \end{aligned} \tag{85}$$

where the moment matrices are given as before, but now the contain an additional spatial transformations, since the regressors are filtered by the inverse spread matrix $\tilde{\mathbf{R}} = \mathbf{R} \otimes \mathbf{I}_n$*:*

$$\mathbf{C}(\mathbf{R}^{-1} \otimes \mathbf{I}_n) vec\mathbf{X}_k = (\mathbf{C}_0 \otimes \mathbf{C}_2) vec\mathbf{X}_k\mathbf{R}'^{-1} = vec\mathbf{C}_0\mathbf{X}_k\mathbf{R}'^{-1}\mathbf{C}'_0 \quad for \quad k = 1, ..., K.$$

The regressor matrix is constructed as

$$\mathbf{X}_{a\rho} = (vec\mathbf{X}_{a\rho,1}, vec\mathbf{X}_{a\rho,2}, ..., vec\mathbf{X}_{a\rho,K})$$

with the elements $vec\mathbf{X}_{a\rho,k} = \mathbf{C}\tilde{\mathbf{R}}^{-1}vec\mathbf{X}_k$ *and the* $K \times K$ *moment matrices (of aggregated filtered regressors) are given by*

$$\mathbf{M}_{\tilde{X}R\tilde{X}} = \left(\begin{pmatrix} vec'\mathbf{X}_{a\rho,1} \\ ... \\ vec'\mathbf{X}_{a\rho,K} \end{pmatrix} (\Omega_a \otimes \mathbf{V}_a)^{-1}\mathbf{X}_{a\rho} \right) = \tag{86}$$

$$\begin{pmatrix} tr\mathbf{X}'_{a\rho,1}\Omega_a^{-1}\mathbf{X}_{a\rho,1}\mathbf{V}_a^{-1} & ... & tr\mathbf{X}'_{a\rho,K}\Omega_a^{-1}\mathbf{X}_{a\rho,1}\mathbf{V}_a^{-1} \\ ... & & ... \\ tr\mathbf{X}'_{a\rho,K}\Omega_a^{-1}\mathbf{X}_{a1}\mathbf{V}_a^{-1} & ... & tr\mathbf{X}'_{aK}\Omega_a^{-1}\mathbf{X}_{a\rho,K}\mathbf{V}_a^{-1}, \end{pmatrix} \tag{87}$$

and for the cross-product moment vector we find

$$\mathbf{M}_{\tilde{X}R\tilde{Y}} = \left(\mathbf{X}'_{a\rho}(\Omega_a \otimes \mathbf{V}_a)^{-1}\mathbf{Y}_a\right) = \begin{pmatrix} tr\mathbf{X}'_{a\rho,1}\Omega_a^{-1}\mathbf{Y}_a\mathbf{V}_a^{-1} \\ ... \\ tr\mathbf{X}'_{a\rho,K}\Omega_a^{-1}\mathbf{Y}_a\mathbf{V}_a^{-1} \end{pmatrix}. \tag{88}$$

Note that these matrices have usually small dimensions, since the number of indicators is limited and can be easily built up by a loop in a computer program.

Note 2. *If we assume a centered g-prior as in (81), then the posterior moments are simply given by*

$$\mathbf{H}_{**}^{-1} = g^{-1}I_K + \sigma^{-2}\mathbf{M}_{\tilde{X}R\tilde{X}}, \quad (89)$$
$$\mathbf{b}_{**} = \sigma^{-2}\mathbf{H}_{**}\mathbf{M}_{\tilde{X}R\tilde{Y}}. \quad (90)$$

(b) *The fcd for the residual variance* σ^{-2} *we find*

$$p(\sigma^{-2} \mid \mathbf{y}, \theta^c) = \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}], \quad (91)$$

with $n_{**} = n_* + n$ *and* $s_{**}^2 n_{**} = s_*^2 n_* + ESS_{\rho,\phi}$ *where the error sum of squares* $ESS_{\rho,\phi}$ *uses aggregated residuals (78) and is given by*

$$ESS_{\rho,\phi} = (\mathbf{y}_a - \mathbf{CR}^{-1}\mathbf{X}\beta)'\mathbf{V}_{a\rho}^{-1}(\mathbf{y}_a - \mathbf{CR}^{-1}\mathbf{X}\beta). \quad (92)$$

Using matrix notation we find

$$ESS_{\rho,\phi} = (vec'\mathbf{U})\mathbf{V}_{a\rho}^{-1}vec\mathbf{U}, \quad (93)$$

with the aggregated residual matrix $\mathbf{U}_a = (\mathbf{u}_1, ..., \mathbf{u}_N) : N \times N$ *is computed by*

$$\mathbf{u}_a = vec\mathbf{U}_a = \mathbf{y}_a - \mathbf{CR}^{-1}\mathbf{X}\beta \quad or \quad \mathbf{u}_i = \mathbf{y}_{a,i} - \mathbf{X}_{a\rho,i}\beta_i, \quad for \quad i = 1, ..., N. \quad (94)$$

(c) *The fcd for the spatial* ρ *coefficient For the generation of the* ρ*'s we use a Metropolis step:*

$$\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau^2] \quad with \quad \alpha = min\left[1, \frac{p(\rho_{new})}{p(\rho_{old}),}\right]$$

the acceptance ratio and where $p(\rho)$ *is the (kernel of) the full conditional for* ρ*, in our case the kernel is just stemming from the likelihood function:*

$$\begin{aligned} p(\rho) &= |\mathbf{V}_{a\rho}|^{-\frac{1}{2}} exp\left(-\frac{1}{2\sigma^2}ESS_{\rho,\phi}\right) \\ &= |\mathbf{R}\Omega^{-1}\mathbf{R}'|^{-\frac{1}{2}} exp\left(-\frac{1}{2\sigma^2}ESS_{\rho,\phi}\right), \end{aligned}$$

because from (78) we find

$$|\mathbf{V}_{a\rho}|^{-\frac{1}{2}} \propto |\mathbf{C}\tilde{\mathbf{R}}^{-1}(\Omega \otimes \mathbf{V})\tilde{\mathbf{R}}'^{-1}\mathbf{C}'|^{-\frac{1}{2}}$$

and $ESS_{\rho,\phi}$ *given in (92) contains* ρ*.*

(d) *The fcd for the correlation parameter* ϕ
For the ϕ *we use a Metropolis step:* $\phi_{new} = \phi_{old} + \mathcal{N}[0, \tau_2^2]$ *with the acceptance ratio* $\alpha = min\left[1, \frac{p(\phi_{new})}{p(\phi_{old})}\right]$ *where* $p(\phi)$ *is the (kernel of) the full conditional for* ϕ*, in our case the kernel is just stemming from the likelihood function:*

$$p(\phi) = |\mathrm{V}\ |^{-\frac{N}{2}} exp\left(-\frac{1}{2\sigma^2}ESS_{\rho,\phi}\right), \quad (95)$$

because from (78) we get

$$|\mathbf{V}_{a\rho}|^{-\frac{1}{2}} \propto |\mathbf{C}\tilde{\mathbf{R}}^{-1}(\Omega \otimes \mathbf{V})\tilde{\mathbf{R}}'^{-1}\mathbf{C}'|^{-\frac{1}{2}}$$

and $ESS_{\rho,\phi}$ given in (92) contains ϕ.

(e) *The fcd for the SUR covariance matrix* $\mathbf{\Omega}$

$$p(\mathbf{\Omega}^{-1} \mid \mathbf{Y}, \Theta^c) = \mathcal{W}_N[\mathbf{\Omega}^{-1} \mid (\nu_{**}\mathbf{\Omega}_{**})^{-1}, \nu_{**}], \tag{96}$$

a (N-dim.) Wishart distribution with $\nu_{**} = \nu_* + N$ d.f. and

$$\mathbf{\Omega}_{**} = \mathbf{\Omega}_* + \mathbf{U}'\mathbf{U}, \tag{97}$$

where $\mathbf{U}$ *is the residual matrix as in (94).*

5.1. The Bayesian origin spatial internal flow (O-SIF) model

We estimate the spatial origin flow (O-SIF or oSIF) model in the same way as the spatial cross-sectional model in Polasek et al. (2009). The model assumes that we have a disaggregated cross-sectional vector $\mathbf{y}_d : n \times 1$ at a certain point in time, which is not observed, but we can observe a shorter, aggregated vector $\mathbf{y}_a = \mathbf{C}\mathbf{y}_d : N \times 1$ and $\mathbf{C}$ is the $N \times n$ aggregation matrix consisting of 0's and 1's, indicating which cells have to be aggregated together. We consider the disaggregated spatial regression model

$$\mathbf{y}_d = \rho \mathbf{W}_d \mathbf{y}_d + \mathbf{X}_d \beta_d + \epsilon_d, \quad \epsilon_d \sim \mathcal{N}[0, \sigma_d^2 \mathbf{I}_n]. \tag{98}$$

The reduced form is obtained by the spread matrix R for an appropriately chosen weight matrix $\mathbf{W}_d : \mathbf{R} = \mathbf{I}_n - \rho_d \mathbf{W}_d$

$$\mathbf{y}_d = \mathbf{R}^{-1}\mathbf{X}_d\beta_d + \mathbf{R}^{-1}\epsilon_d, \quad \mathbf{R}^{-1}\epsilon_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2(\mathbf{R}'\mathbf{R})^{-1}]. \tag{99}$$

The prior distribution for the parameters $\theta_d = (\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d)$ is proportional to

$$\begin{aligned} p(\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d) &\propto p(\beta_d) \cdot p(\sigma_d^{-2}) \\ &= \mathcal{N}[\beta_d \mid \boldsymbol{\beta}_*, \mathbf{H}_*] \cdot \Gamma(\sigma_d^{-2} \mid s_*^2, n_*), \end{aligned} \tag{100}$$

since we assume a uniform prior for $\rho_a \sim \mathcal{U}[-1, 1]$. The C-aggregation of the reduced form model is obtained by multiplying with the $N \times n$ matrix C

$$Cy_d = C\mathbf{R}^{-1}X_d\beta_d + C\mathbf{R}^{-1}\epsilon_d, \quad C\mathbf{R}^{-1}\epsilon_d \sim \mathcal{N}[\mathbf{0}, \sigma_d^2 C(\mathbf{R}'\mathbf{R})^{-1}C']. \tag{101}$$

We will write shorter for the covariance matrix:

$$\sigma_d^2 \Sigma(\rho_d) = \sigma_d^2 C (\mathbf{R}_d' \mathbf{R}_d)^{-1} C'. \tag{102}$$

We see that (101) is a completely observed model for the disaggregated parameters θ_d but estimated with an aggregated $y_a = C y_d$ variable. The joint distribution of $\theta_d = (\beta_d, \rho_d, \sigma_d^2)$ and the aggregated data y_a of this oSIF model is given by

$$p(\theta_d, y_a) = \mathcal{N}[\mathbf{C}\mathbf{R}^{-1} X_d \beta_d, \sigma_d^2 \Sigma(\rho_d)] \cdot \mathcal{N}[\beta_a \mid \boldsymbol{\beta}_*, \mathbf{H}_*] \cdot \Gamma[\sigma_d^{-2} \mid s_*^2, n_*]. \tag{103}$$

Note that the parameters of the aggregated model θ_a can be estimated in the RF model for the aggregated data. Since the data set and the model differ, the results of the estimation are different.

$$y_a = \mathbf{R}^{-1} X_a \beta_a + \mathbf{R}^{-1} \epsilon_a, \quad \mathbf{R}_a^{-1} \epsilon_a \sim \mathcal{N}[0, \Sigma_a = \sigma_a^2 (\mathbf{R}'\mathbf{R})^{-1}]. \tag{104}$$

The prior distribution for the parameters $\theta_a = (\boldsymbol{\beta}_a, \sigma_a^{-2}, \rho_a)$ is proportional to

$$\begin{aligned} p(\boldsymbol{\beta}_a, \sigma_a^{-2}, \rho_a) &\propto p(\beta_d) \cdot p(\sigma_a^{-2}) \\ &= \mathcal{N}[\beta_a \mid \boldsymbol{\beta}_*, \mathbf{H}_*] \cdot \Gamma(\sigma_a^{-2} \mid s_*^2, n_*), \end{aligned}$$

since we have assumed a uniform prior for $\rho_a \sim U[-1, 1]$.

The MCMC for this model follows the same steps as in Theorem (5) but now with the parameters θ_a since the joint distribution is

$$p(\theta_a, \mathbf{y}_a) = \mathcal{N}[\mathbf{R}_a^{-1} \mathbf{X}_a \beta_a, \sigma_a^2 \Sigma(\rho_a)] \cdot \mathcal{N}[\beta_a \mid \boldsymbol{\beta}_*, \mathbf{H}_*] \cdot \Gamma(\sigma_a^{-2} \mid s_*^2, n_*). \tag{105}$$

We will write shorter for the covariance matrix:

$$\sigma_a^2 \Sigma(\rho_a) = \sigma_a^2 (\mathbf{R}_a' \mathbf{R}_a)^{-1}.$$

The prior distribution for the parameters $\theta_d = (\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d)$ is proportional to

$$\begin{aligned} p(\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d) &\propto p(\beta_a) \cdot p(\sigma_d^{-2}) \\ &= \mathcal{N}[\beta_d \mid \boldsymbol{\beta}_*, \mathbf{H}_*] \cdot \Gamma(\sigma_d^{-2} \mid s_*^2, n_*), \end{aligned}$$

since we assume a uniform prior for $\rho_d \sim \mathcal{U}[-1, 1]$.

In case of the heteroskedastic model we have the parameter vector $\Theta_d = (\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d, \Omega_d, \mathbf{V}_d)$ and the prior distribution has to be enlarged by 2 Wishart distributions:

$$\begin{aligned} p(\Theta_d) &\propto p(\boldsymbol{\beta}_d, \sigma_d^{-2}, \rho_d) p(\Omega_d) p(\mathbf{V}_d) \\ &= p(\theta_d) \cdot \mathcal{W}[\Omega_d^{-1} \mid (\Omega_* \kappa_*)^{-1}, \kappa_*] \mathcal{W}[\mathbf{V}_d^{-1} \mid (\mathbf{V}_* \nu_*)^{-1}, \nu_*]. \end{aligned}$$

5.2. MCMC for the oSIF Chow-Lin model

After vectorizing the flow matrices we obtain a cross-sectional Chow-Lin SIF model (oSIF) for the parameters θ_d given in (101) and let us denote the 3 conditional distributions by $p(\rho_d \mid \theta^c), p(\beta_d \mid \theta^c)$, and $p(\sigma_d^2 \mid \theta^c)$ where $\theta_d = (\rho_d, \beta_d, \sigma_d^2)$ denotes all the parameter of the model and θ^c the complementary parameters in the f.c.d.'s, respectively. The MCMC procedure consists of 3 blocks of sampling, as is shown in the next theorem:

Theorem 6 (MCMC in the homoskedastic oSIF model). *The MCMC estimation for the oSIF Chow-Lin model involves the following iterations:*

Step 1. *Draw $\boldsymbol{\beta}_d$ from $\mathcal{N}\left[\boldsymbol{\beta}_a \mid \mathbf{b}_{**}, \mathbf{H}_{**}\right]$*
Step 2. *Draw ρ_i by a Metropolis step: $\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau^2]$*
Step 3. *Draw σ_d^{-2} from $\Gamma[\sigma_d^{-2} \mid s_{**}^2, n_{**}]$*
Step 4. *Repeat until convergence.*

Proof 6 (Proof of Theorem 6).

(a) *The fcd for the beta regression coefficients is*

$$\begin{aligned} p(\boldsymbol{\beta}_d \mid \mathbf{y}_a, \theta^c) &= \mathcal{N}[\boldsymbol{\beta}_d | \mathbf{b}_*, \mathbf{H}_*] \cdot \\ & \quad \mathcal{N}[\mathbf{y}_a \mid \mathbf{C}\mathbf{R}^{-1}\mathbf{X}_d\beta_d, \sigma_d^2\mathbf{C}(\mathbf{R}'\mathbf{R})^{-1}C'] \\ &= \mathcal{N}\left[\boldsymbol{\beta}_d \mid \mathbf{b}_{**}, \mathbf{H}_{**}\right], \end{aligned} \tag{106}$$

with the parameters

$$\begin{aligned} \mathbf{H}_{**}^{-1} &= \mathbf{H}_*^{-1}\mathbf{b}_* + \sigma_d^{-2}\mathbf{X}'\mathbf{R}'^{-1}\mathbf{C}'\Sigma(\rho_d)^{-1}\mathbf{C}\mathbf{R}^{-1}\mathbf{X}, \\ \mathbf{b}_{**} &= \mathbf{H}_{**}[\mathbf{H}_*^{-1}\mathbf{b}_* + \sigma_d^{-2}\mathbf{X}'\mathbf{R}'^{-1}\mathbf{C}'\Sigma(\rho_d)^{-1}\mathbf{y}_a]. \end{aligned}$$

(b) *For the fcd for the inverse variance we find*

$$p(\sigma_d^{-2} \mid \mathbf{y}_a, \theta^c) = \Gamma[\sigma_d^{-2} \mid s_{**}^2, n_{**}], \tag{107}$$

with $n_{**} = n_* + n$ and $s_{**}^2 n_{**} = s_*^2 n_* + ESS_{\rho_d}$ and where the error sum of squares ESS_{ρ_d} is given by

$$ESS_\rho = (\mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}\mathbf{X}_d\beta_a)'\Sigma(\rho)^{-1}(\mathbf{y}_a - \mathbf{C}\mathbf{R}^{-1}\mathbf{X}_d\beta_a). \tag{108}$$

(c) *For the fcd of the spatial ρ_d we use a Metropolis step:*

$$\rho_{new} = \rho_{old} + \mathcal{N}(0, \tau^2) \quad \text{with } \alpha = min\left[1, \frac{p(\rho_{new})}{p(\rho_{old})}\right]$$

being the acceptance ratio and where $p(\rho_d)$ is the (kernel of) the full conditional for ρ_d, in our case the kernel is just stemming from the likelihood function:

$$p(\rho_d) = |\Sigma(\rho_d)|^{-\frac{1}{2}} exp(-\frac{1}{\sigma_d^2} ESS_{\rho_d}),$$

with ESS_{ρ_d} given in (92).

From the MCMC simulation run we obtain a numerical sample of the posterior distribution $p(\beta_d, \rho_d, \sigma_d^{-2} \mid \mathbf{y}_a)$.

Theorem 7 (MCMC in the heteroskedastic oSIF model). *The MCMC estimation for the oSIF Chow-Lin model involves the following iterations:*

Step 1. *Draw* $\boldsymbol{\beta}_a$ *from* $\mathcal{N}[\boldsymbol{\beta}_a \mid \mathbf{b}_{**}, \mathbf{H}_{**}]$
Step 2. *Draw* ρ_i *by a Metropolis step:* $\rho_{new} = \rho_{old} + \mathcal{N}(0, \tau^2)$
Step 3. *Draw* σ_d^{-2} *from* $\Gamma[\sigma_a^{-2} \mid s_{**}^2, n_{**}]$
Step 4. *Draw* Ω_d^{-1} *from* $\mathcal{W}[\Omega_d^{-1} \mid (\Omega_{**}\kappa_{**})^{-1}, \kappa_{**}]$
Step 5. *Draw* $\mathbf{v}_d^{-1}$ *from* $\mathcal{W}[\mathbf{v}_d^{-1} \mid (\mathbf{v}_{**}\nu_{**})^{-1}, \nu_{**}]$
Step 6. *Repeat until convergence.*

Proof 7 (Proof of Theorem 7). *Add to the iteration cycle of the homoskedastic oSIF model the following 2 draws*

1. *The fcd for the precision matrix* Ω_d^{-1}. *The posterior d.f. are given by* $\kappa_{**} = \kappa_* + n$ *and the scale matrix by*

$$\Omega_{**}\kappa_{**} = \Omega_*\kappa_* + \hat{U}\hat{U}'$$

with $\hat{U} = \mathbf{Y}_a - \hat{Y}_a$ *and the matrix* $\hat{Y}$ *contains the current predicted values of the RF.*

2. *The fcd for the precision matrix* $\mathbf{V}_d^{-1}$. *The posterior d.f. are given by* $\nu_{**} = \nu_* + n$ *and the scale matrix by* $\mathbf{V}_{**}\nu_{**} = \mathbf{V}_*\nu_* + \hat{U}'\hat{U}$ *with* $\hat{U} = \mathbf{Y}_a - \hat{Y}_a$ *as before.*

6. Application to trade flows in European regions

The following section demonstrates the method of the origin-destination Chow-Lin procedure. We predict the regional NUTS-2 origin/destination import flows for a sample of 10 EU-27 countries by a classical trade gravity model (see for example Frankel and Romer (1999)). The aggregate trade data for the year 2006 is taken from the Eurostat external trade database. As explaining factors we use origin and destination population and GDP as well as the Euclidean distances between the countries and regions. The regional GDP and population data was obtained from Eurostat.

The aggregate trade flows are first estimated using the OLS estimation algorithm of LeSage (1998). We use the following gravity specification for the estimation of the trade flows:

$$\begin{aligned} \log Y_{ij} &= \alpha + \beta_1 \log GDP_i + \beta_2 \log GDP_j \\ &\quad + \gamma_1 \log POP_i + \gamma_2 \log POP_j + \delta \log Dist_{ij} + \epsilon_{ij}, \end{aligned} \tag{109}$$

with Y_{ij} being the trade flow from region i to region j, GDP_i (GDP_j) is the Gross Domestic Product of the origin (destination) region, POP is the population and $Dist_{ij}$ is the Euclidean distance between region i and region j. The results of this estimation are given in Table 1. The signs of the coefficients are as expected by theory. GDP and population of the destination region act as an attraction factor for trade flows. Conversely, the population of the origin acts as a proxy for the demand of the exporting regions. The more is demanded locally, the less is exported. GDP of the origin region can be interpreted as a supply factor for potential exports. Distance serves as a proxy for transport costs of trade and has the expected negative impact. All coefficients are highly significant and the R^2 is of sufficient size.

Table 1: OLS Chow-Lin flow estimation results

variable	coef	std. err.	
constant	-1.461	2.402	
GDP origin	2.124	0.237	***
GDP destination	0.753	0.237	***
POP origin	-1.248	0.214	***
POP destination	0.152	0.214	
Distance	-1.116	0.135	***
adj. R^2		0.79	
Nobs		90	

Using the non-spatial Chow-Lin flow predictor, the import flows have been disaggregated as shown in Figure 1. As the sum of the logs is not equal to the log of the sum we notice that the Chow-Lin summation property, i.e. that the sum of the disaggregates equals the aggregates, does not hold for the exponentiated (antilog) values of the flows.

The lowest flows are from the Spanish autonomous region Melilla in North Africa (with region number 42), the highest import flows are observed for the Greek region Attiki (Athens).

Embedding the Chow-Lin procedure in the spatial framework, we define the spatial weight matrix $\mathbf{W}$ as the row-normalized inverse Euclidean distances between the countries or regions. The results of the aggregate D-SIF (i.e. the $W_{destination}$-matrix has been constructed by $\mathbf{I}_N \otimes \mathbf{W}$) estimation are given in Table 2. We receive a significant spatial effect and the size of the remaining coefficients is reduced.

Using the SAR GLS Chow-Lin flow predictor, the import flows can be disaggregated as shown in Figure 2 in form of a matrix contour plot. The predicted flow pattern is rather similar to the non-spatial. As no information is available on observed regional trade flows, no evaluation measures can be computed so far.

Figure 1: Predicted log import flows, 2006

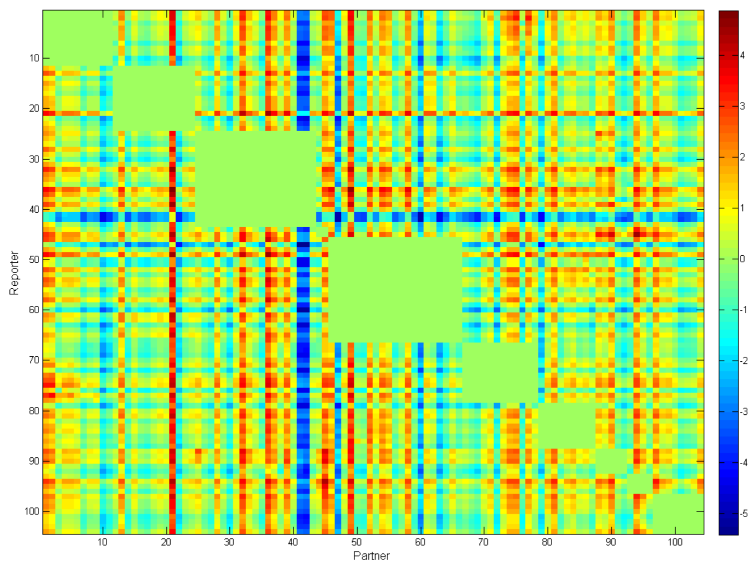
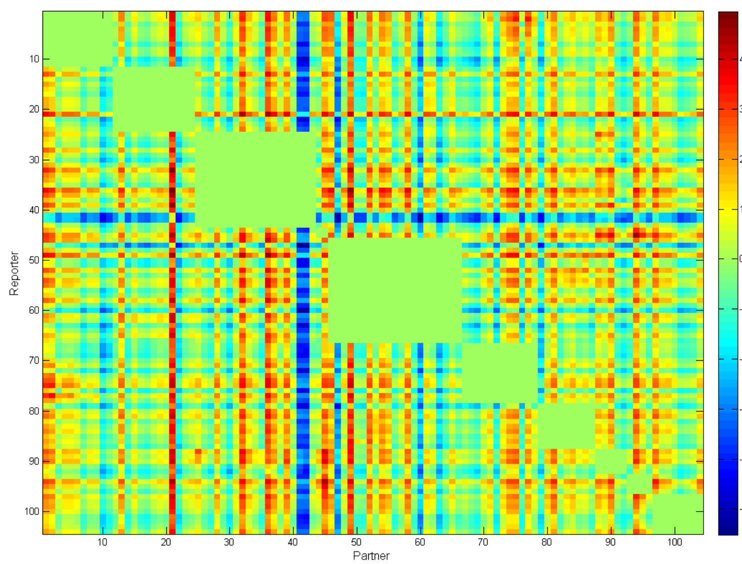


Table 2: SAR Chow-Lin flow estimation results

variable	coef	std. err.	
constant	-6.109	2.187	***
GDP origin	1.040	0.279	***
GDP destination	0.858	0.200	***
POP origin	-0.596	0.213	***
POP destination	0.176	0.179	
Distance	-1.005	0.118	***
Spatial Lag	0.481	0.086	***
adj. R^2		0.77	
Nobs		90	

Figure 2: Predicted NUTS-2 log import flows, 2006 - SAR



7. Conclusions

This paper has shown that the Chow-Lin method can be applied in a classical or Bayesian framework for completing data in spatial internal flow (SIF) models, i.e. for origin and destination (O-D) data. The spatial models rely on a spatial lag polynomial that takes into account the neighborhood relationships at all origins and destinations. The Bayesian approach is based on MCMC and has the advantage that the problem of missing data can be easily implemented in quite flexible MCMC programs. At the end of the paper, we demonstrated the non-spatial and spatial internal flow model with an application to European trade flow data within a gravity framework. Unfortunately, no regional observed trade flow data is available, so we are not able to evaluate the accuracy of the forecasts.

In further research, we will consider models where we can observe external flows, i.e. flows from the current N aggregate units to other units outside the N current units where we know only the aggregated sum of the flows. Such models are called spatial external flows (SEF) models and details will be presented in an upcoming paper. The basic ideas for spatial Chow-Lin (SCL) models were discussed in Polasek and Sellner (2008).

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Chow, G. C. and Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375.

Frankel, J. A. and Romer, D. (1999). Does trade cause growth? *The American Economic Review*, 89(3):379–399.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *American Statistical Association Journal*, 57:369–375.

LeSage, J. P. (1998). Spatial econometrics. Technical report, University of Toledo.

LeSage, J. P. and Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

Polasek, W. and Sellner, R. (2008). Spatial chow-lin methods: Bayesian and ml forecast comparison. Working Paper 38-08, The Rimini Centre for Economic Analysis.

Polasek, W., Sellner, R., and Llano, C. (2009). Bayesian methods for completing space-time panel models. Economics Series 241, Institute for Advanced Studies.

Authors: Wolfgang Polasek, Richard Sellner

Title: Spatial Chow-Lin Methods for Data Completion in Econometric Flow Models

Reihe Ökonomie / Economics Series 255

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996